Title (en)
METHOD AND SYSTEM FOR DETECTING AN ACTION IN A VIDEO CLIP

Title (de)
VERFAHREN UND SYSTEM ZUR ERKENNUNG EINER HANDLUNG IN EINEM VIDEOCLIP

Title (fr)
PROCÉDÉ ET SYSTÈME POUR DÉTECTER UNE ACTION DANS UN CLIP VIDÉO

Publication
EP 3995992 A1 20220511 (EN)

Application
EP 20306343 A 20201106

Priority
EP 20306343 A 20201106

Abstract (en)
A method for detecting an action in a video clip comprising a succession of video frames (12), comprising:- obtaining spatio-temporal features (F_0) extracted from the video clip;- inputting the spatio-temporal features (F_0) to a first artificial neural network (20) having a first temporal receptive field;- for i from 2 to N, N greater than or equal to 2, inputting the spatio-temporal features (F_0) and the output (F_i) of the i-1th artificial neural network to a ith artificial neural network having a ith temporal receptive field different from the i-1th temporal receptive field ;- determining whether a given action is represented in the video clip based on the output of the Nth artificial neural network;wherein at least the jth artificial neural network (20), where j lies between 1 and N, comprises a self-attention mechanism.

IPC 8 full level
G06K 9/00 (2022.01); G06N 3/04 (2006.01)

CPC (source: EP)
G06N 3/045 (2023.01); G06V 10/454 (2022.01); G06V 20/41 (2022.01)

Citation (applicant)
- VASWANI, A.SHAZEER, N.PARMAR, N.USZKOREIT, J.,JONES, L.GOMEZ, A.N.,KAISER, L.POLOSUKHIN, I.: "Attention is All you Need", ARXIV, ABS/1706.03762, 2017
- SERENA YEUNGOLGA RUSSAKOVSKYNING JINMYKHAYLO ANDRILUKAGREG MORILI FEI-FEI: "Every moment counts: Dense detailed labeling of actions in complex videos", INTERNATIONAL JOURNAL OF COMPUTER VISION, vol. 126, no. 2-4, 2018, pages 375 - 389
- GUNNAR A. SIGURDSSONG"UL VAROLXIAOLONG WANGALI FARHADIIVAN LAPTEVABHINAV GUPTA: "Hollywood in homes: Crowdsourcing data collection for activity understanding", EUROPEAN CONFERENCE ON COMPUTER VISION (ECCV), 2016
- AJ PIERGIOVANNIMICHAEL S. RYOO: "Temporal Gaussian mixture layer for videos", INTERNATIONAL CONFERENCE ON MACHINE LEARNING (ICML), 2019

Citation (search report)
- [A] CN 111563404 A 20200821 - UNIV BEIJING
- [XI] XU BAOHAN ET AL: "Dense Dilated Network for Video Action Recognition", IEEE TRANSACTIONS ON IMAGE PROCESSING, IEE SERVICE CENTER , PISCATAWAY , NJ, US, vol. 28, no. 10, 1 October 2019 (2019-10-01), pages 4941 - 4953, XP011738227, ISSN: 1057-7149, [retrieved on 20190801], DOI: 10.1109/TIP.2019.2917283
- [I] COLIN LEA ET AL: "Temporal Convolutional Networks for Action Segmentation and Detection", ARXIV.ORG, CORNELL UNIVERSITY LIBRARY, 201 OLIN LIBRARY CORNELL UNIVERSITY ITHACA, NY 14853, 16 November 2016 (2016-11-16), XP080732197, DOI: 10.1109/CVPR.2017.113
- [I] ZHAI MINGLIANG ET AL: "Optical Flow Estimation Using Dual Self-Attention Pyramid Networks", IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, INSTITUTE OF ELECTRICAL AND ELECTRONICS ENGINEERS, US, vol. 30, no. 10, 21 September 2019 (2019-09-21), pages 3663 - 3674, XP011812440, ISSN: 1051-8215, [retrieved on 20201001], DOI: 10.1109/TCSVT.2019.2943140
- [I] ZHAI MINGLIANG ET AL: "Optical flow estimation using channel attention mechanism and dilated convolutional neural networks", NEUROCOMPUTING, ELSEVIER, AMSTERDAM, NL, vol. 368, 27 August 2019 (2019-08-27), pages 124 - 132, XP085859009, ISSN: 0925-2312, [retrieved on 20190827], DOI: 10.1016/J.NEUCOM.2019.08.040

Cited by
CN114842559A; CN115049945A

Designated contracting state (EPC)
AL AT BE BG CH CY CZ DE DK EE ES FI FR GB GR HR HU IE IS IT LI LT LU LV MC MK MT NL NO PL PT RO RS SE SI SK SM TR

Designated extension state (EPC)
BA ME

DOCDB simple family (publication)
EP 3995992 A1 20220511

DOCDB simple family (application)
EP 20306343 A 20201106